Dynamic Patterns of Academic Forum Activities

Zhi-Dan Zhao[1,2], Ya-Chun Gao[3], Shi-Min Cai[1,2],[*] and Tao Zhou[1,4][†]

1 *CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China*
2 *State Key Laboratory of Networking and Switching Technology (Beijing*
University of Posts and Telecommunications), Beijing, 100876, P. R. China
3 *School of Management and Economics, University of Electronic*
Science and Technology of China, Chengdu, 611731, P. R. China and
4 *Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China*

A mass of traces of human activities show rich dynamic patterns. In this article, we comprehensively investigate the dynamic patterns of 50 thousands of researchers' activities in *Sciencenet*, the largest multi-disciplinary academic community in China. Through statistical analyses, we found that (i) there exists a power-law scaling between the frequency of visits to an academic forum and the number of corresponding visitors, with the exponent being about 1.33; (ii) the expansion process of academic forums obeys the Heaps' law, namely the number of distinct visited forums to the number of visits grows in a power-law form with exponent being about 0.54; (iii) the probability distributions of time intervals and the number of visits taken to revisit the same academic forum both follow power-laws, indicating the existence of memory effect in academic forum activities. On the basis of these empirical results, we propose a dynamic model that incorporates the exploration, preferential return with memory effect, which can well reproduce the observed scaling laws.

PACS numbers: 89.75.Da, 89.65.-s, 02.50.-r

I. INTRODUCTION

It has been deemed that complicated factors are affecting the dynamic patterns of human activities, such as the priority of task [1–4], individual interest [5–7], memory effects [8–13], deadline effects [14], social contacts [15–18], and so on. Relevant practical applications range from information spreading [19–22], decision making [23] to advertising [24, 25] and recommendation [26–28].

Although large amounts of empirical results on human dynamics have been reported in various fields [15, 29–32], researchers' activities in academic forums are rarely investigated and still not clearly understood as they usually accompanies with many endogenous and exogenous factors, including the individual preference and professional background, the quality of a forum, the content of a post, and so on. To fill this gap, we study a data set sampled from *Sciencenet* (http://www.sciencenet.cn/) that contains the academic forum activities of many Chinese researchers. We observe novel dynamic patterns characterized by the following statistical features: (i) the power-law relation between the frequency of visits to an academic forum and the number of corresponding visitors; (ii) the Heaps' law [33] in the expansion process; (iii) the memory effect. We further propose a dynamic model that well reproduces the empirical observations.

II. EMPIRICAL RESULTS

Sciencenet is the largest multi-disciplinary academic community in China, which contains a blog system, a bulletin board system (BBS, consisting of 60 academic forums), and a virtual social network of researchers. Our data set keeps track of activities in the BBS between October/1/2007 and July/7/2011, composed of 366,524 records from 49,578 researchers. Each record includes researcher ID, academic forum ID, posting/reviewing topic ID, and timestamp with resolution of minute. Table I presents the names, the visiting frequencies (i.e, total visits from all researchers), and the number of visitors of the 60 academic forums, ranked in the descending order of visiting frequencies.

Rank	Forum Name	Frequency	#visitors
1	Mathematics	28,218	5,849
2	Materials Science	23,610	6,293
3	Geology and Geophysics	21,363	4,025
4	Oceanography	21,265	4,019
5	Reading	18,586	5,315
6	Research Experience	15,757	6,713
7	Tea Break	13,351	4,209
8	Fund Application	11,672	4,045
9	Article	11,511	4,983
10	Analytical Chemistry	9,228	2,769
11	High-polymer Chemical	8,271	2,536
12	Nanotechnology	8,211	2,815
13	Amorphous Alloy and Glass	8,148	1,151
14	Computer Science	7,834	2,709
15	Chemical Engineering	7,541	2,208
16	Catalysts	7,541	2,208
17	Organic	6,952	2,210

[*]Electronic address: shimin.cai81@gmail.com
[†]Electronic address: zhutou@ustc.edu

18	Physics	6,702	2,547
19	Chemistry	6,653	2,860
20	Physical Resources	6,289	2,402
21	References	6,102	3,115
22	Study Abroad	6,085	2,781
23	Optics	5,984	1,759
24	Molecular and Cellular Immunology	5,544	2,132
25	Electronic Power	5,539	1,841
26	Mechanics	5,306	1,736
27	Control of Intelligent Modeling	5,044	1,763
28	Electrochemistry	4,961	1,453
29	Crystal	4,775	1,881
30	Mechanical	4,772	1,483
31	Theoretical Physics	4,732	1,804
32	Communication	4,691	1,600
33	Civil Engineering	4,164	905
34	Botany and Zoology	3,973	1,577
35	Clinical Medicine	3,149	1,021
36	Condensed Matter	3,130	1,195
37	Examination	2,849	1,460
38	Resources and Environment	2,648	1,271
39	Fine Chemicals	2,464	772
40	Materials and Methods	2,452	1,397
41	Pharmaceutical Chemistry	2,357	677
42	Chinese Medicine	2,170	548
43	History of Science	2,142	991
44	Engineering	2,040	1,080
45	Management	1,920	978
46	Academic Exhibition	1,920	978
47	GPS/RS/GIS	1,862	757
48	Genetics	1,776	769
49	Agronomy	1,776	769
50	General Chemistry	1,692	824
51	Astrophysics	1,565	509
52	Biophysical and Bioinformatics	1,485	711
53	Molecular Simulation	1312	571
54	Energy Science	1,233	546
55	Brain Science	1,161	553
56	Petroleum Exploration	1,089	450
57	Others	631	500
58	Cryptography	625	251
59	Alloy	427	176
60	Medical and Health Management	275	196

TABLE I: Basic information of the 60 academic forums.

At the aggregated level, as shown in Table I, both the number of visits and the number of visitors are heterogeneously distributed among forums. While at the individual level, the visiting behavior is also heterogeneous, indicated by the burstiness that a visitor usually stays in a forum for long time and then glances over several forums. Figure 1 illustrates the transition process of two example visitors, with visiting sequences shown in the first rows of Fig. 1(c) and 1(d). Figure 2 shows the transition process of a typical real visitor, with two different
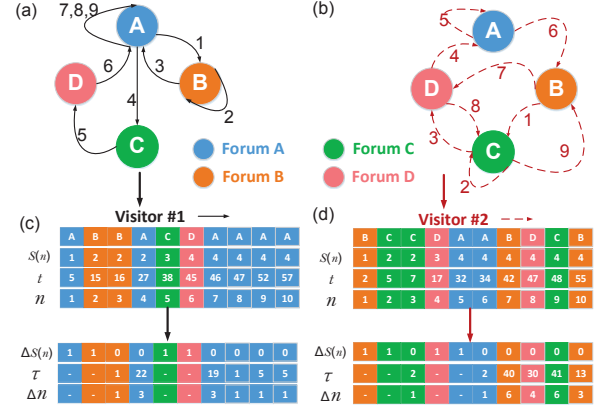

FIG. 1: (Color online) An illustration of concepts in view of two visitors' activities. (a)-(b) The transition process among different academic forums (black solid and red dash lines represent visitor #1 and visitor #2, respectively). (c)-(d) The illustration of visiting sequences, as well as other relevant parameters of the two visitors, including the number of distinct academic forums $S(n)$, the real time t, the number of visits n, the exploration of new academic forums $\Delta S(n)$, the real time interval τ and the click time interval Δn.

scales: (a) real time and (b) click time. The colored vertical lines represent different academic forums.
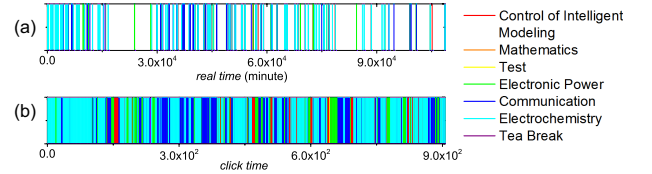

FIG. 2: (Color online) A typical visitor's transition process on academic forums along with (a) real time and (b) click time, suggesting a heterogeneous dynamic pattern that commonly exists in other online human dynamics [32]. The colored vertical line denotes the different academic forums.

As shown in Fig. 3, there is a superlinear correlation between the number of visits to a academic forum (denoted by F) and the number of corresponding visitors (denoted by P), as $F \sim P^{\gamma}$ with $\gamma \approx 1.33$. This superlinear relationship maybe resulted from two reasons. Firstly, the forum that attracts more visitors is usually of higher quality, namely a post in this forum can attract a higher fraction of visitors in average. Secondly, many visits are induced by some previously comments and replies [34], in particular, visitors often care much about the replies to their own posts and comments. Such social cascading process may lead to a superlinear growing trend since the maximum possible volume of social interaction is of the order P^2.

The expansion process of a visitor can be described by the number of distinct visited forums $S(n)$, where n is the number of visits of this visitor. Taking visi-
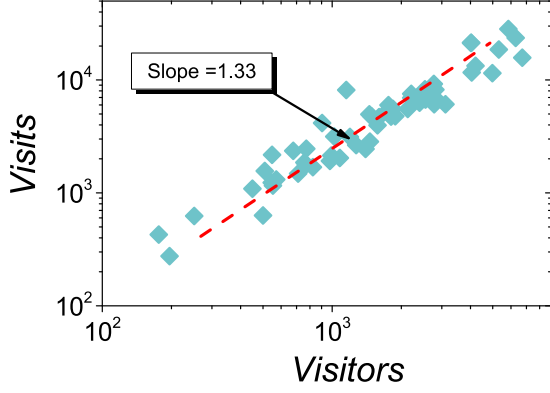
FIG. 3: (Color online) A power-law relation between the frequency of visits to an academic forum and the number of corresponding visitors. Each data point stands for an academic forum.
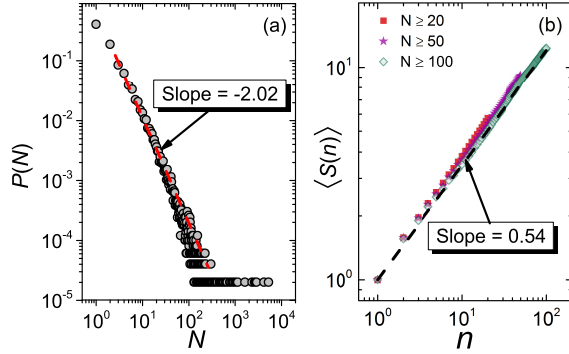

FIG. 4: (Color online) (a) The probability distribution of the number of visits N for all visitors. (b) The scaling behavior in the expansion process. The black dash line indicates a power-law, $S \sim n^\alpha$, with scaling exponent $\alpha \approx 0.54$. The red squares, purple stars and green diamonds represent three groups of visitors with more than 20, 50 and 100 visits, respectively.

tor #1 in Fig. 1 as an example, his visiting sequence is {A,B,B,A,C,D,A,A,A,A}, and the corresponding $S(n)$ sequence is $\{1, 2, 2, 2, 3, 4, 4, 4, 4, 4\}$, as shown in the second row of Fig. 1(c). The activity level is a key feature of an user [32], which may affect the expansion process. Here we use the total number of visits N_i to quantify the activity level of user u_i. As shown in Fig. 4(a), the distribution of users' activity levels is very heterogeneous, following a power law with exponent about 2.02, in accordance with the empirical results in other online activities [32]. In common sense, users with different activity levels may behave differently, hence we pick out three groups of visitors whose visits are more than 20, 50 and 100 times, respectively. Their expansion sequence $S(n)$ in respect to the number of visits n is presented in

Fig. 4(b), suggesting a robust scaling behavior $S \sim n^\alpha$, with the power-law exponent $\alpha \approx 0.54$, similar to some observations in human mobility [35, 36].

To see the preference to visit a new forum, we look at the exploration sequence $\Delta S(n) = S(n+1) - S(n)$. Simple examples about the calculation of $\Delta S(n)$ are shown in Fig. 1(c) and 1(d). As shown in Fig. 5(a), the rescaled exploration $\frac{\Delta S(n)}{\langle \Delta S(n) \rangle}$ scales in a power-law form with the number of visits, as $\frac{\Delta S(n)}{\langle \Delta S(n) \rangle} \sim n^{-\beta}$, where $\beta \approx 0.50$ and the rescaling factor $\langle \Delta S(n) \rangle$ is averaged over all visitors in the considered group. The result confirms the scaling behavior found in expansion process since $\alpha + \beta \approx 1$. Given the target visitor, if we denote P_{new} the probability to visit a new forum and $1 - P_{new}$ the probability to return to a previously visited forum, then the above result suggests a scaling relation $P_{new} \sim pn^{-\beta}$, where p is a constant. Figure 5(b) presents the distributions of p for the three groups of visitors, which are bell-shaped curves with mean values as $\langle p \rangle \approx 0.41$, $\langle p \rangle \approx 0.36$ and $\langle p \rangle \approx 0.32$ for $N \geq 20$, $N \geq 50$ and $N \geq 100$, respectively.
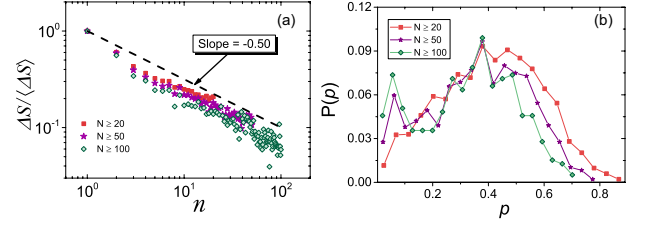

FIG. 5: (Color online) (a) The scaling behavior of rescaled exploration $\frac{\Delta S(n)}{\langle \Delta S(n) \rangle}$ in respect to the number of visits n. (b) The probability distributions of the values of p for individual visitors. The definition of groups is same to that of figure 4.

Memory is a very significant feature of human activities [5, 7–13, 32, 37], which can be characterized by the probability distribution of the real time interval τ between two consecutive visits to the same forum or the number of visits Δn taken to revisit the same forum. For example, Δn for the forum B in the visiting sequence of visitor #2 are 6 and 3, as shown in Fig. 1(d). More examples for both τ and Δn can be found in Fig. 1(c) and 1(d). Figure 6 reports the probability distributions of τ and Δn of the three groups of visitors. Both $P(\tau)$ and $P(\Delta n)$ can be well fitted by power laws with exponents being 1.88 and 1.89, respectively. Such distributions suggest the existence of the memory effect since a visitor has higher probability to return to recently visited forums.

III. MODEL

The above-mentioned empirical results provide us insights to the dynamic patterns of visiting to academic forums. In particular, the memory effect assigns higher probability to a visitor to return to the forums being
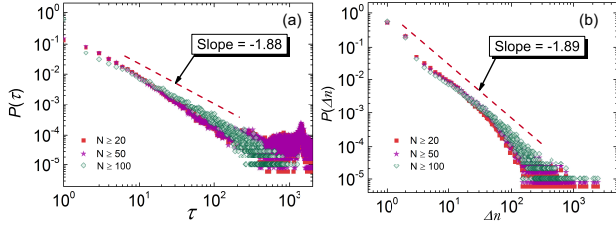
FIG. 6: (Color online) The memory effect in visiting activities, suggested by the power-law distributions of real time interval τ and (b) the number of visits Δn. The guiding curve follows a power-law form, $\tau^{-\gamma}$ and $\Delta n^{-\xi}$ with $\gamma \approx 1.88$ and $\xi \approx 1.89$, obtained by the maximum likelihood method [38]. The definition of groups is same to that of figure 4.
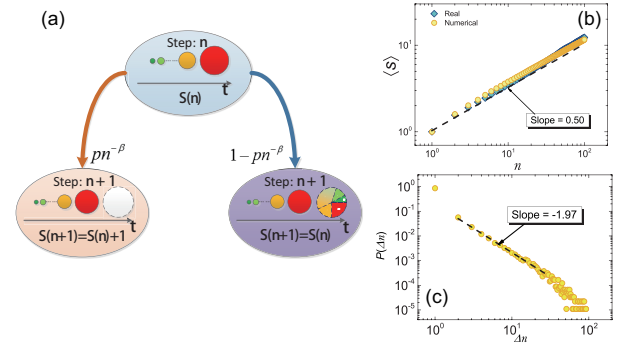

FIG. 7: (Color online) (a) The schematic of the individual dynamic model. (b) The scaling behaviors of expansion $S(n)$ in respect to n for empirical data and the model. (c) The power-law distribution of return interval Δn obtained from the model. The parameters are set as $N = 100$, $\langle p \rangle = 0.41$, $\beta = 0.50$, $\xi = 1.89$, and the number of visitor is 10^3, namely all the results can be considered as being averaged over 10^3 independent runs.

visited recently. Accordingly, we propose a model incorporating three generic ingredients: (i) the exploration, indicating the tendency to visit a new forum; (ii) the preferential return, indicating the tendency to visit a previously visited forum, with frequently visited ones being more attractive; (iii) the memory effect, indicating the tendency to revisit a recently visited forum. Notice that, the preferential return can be mathematically considered as a kind of memory with infinite memory length, however, we distinguish it from memory effect since the preferential return may be resulted from the quality of a forum and the match between the forum's content and the visitor's professional background.

Figure 7(a) illustrates a schematic for the individual dynamic model. More specifically, after the nth visit, the target visitor has two choices for the next visit: (i) to visit a new academic forum with probability P_{new} that depends on the number of previous visits n; (ii) to revisit one of the $S(n)$ previously visited forums with the complementary probability $(1 - P_{new})$. If the visitor visits a new forum at the $(n + 1)$th visit, the expansion sequence updates as $S(n + 1) = S(n) + 1$, otherwise it keeps $S(n + 1) = S(n)$. Initially, $S(1) = 1$, and the probability P_{new} evolves as

$$P_{new} \propto \frac{dS}{dn} = pn^{-\beta}. \quad (1)$$

In our model, the parameter $\beta = 0.50$ is obtained from the empirical scaling exponent for the case $N \geq 20$ (see Fig. 5(a)), while the parameter $\langle p \rangle = 0.41$ is estimated from the mean value of p for the case $N \geq 20$ (see Fig. 5(b)).

When considering the return to a visited forum, we combine preferential return and memory effect into a unified probability. The preferential return indicates that every visit to a forum will contribute to the probability to revisit this forum, while the memory effect suggests a power-law decay of the contribution strength. Given a forum i, if it appears f_i times in the target visitor's visiting sequence, we denote the discrete time interval between the jth appearance to the current time step $(n+1)$

as

$$\Delta n_j^{(i)} = n + 1 - T_j^{(i)}, \quad (2)$$

where $T_j^{(i)}$ is the very time step when i just appears j times. As discussed above, every visit will be counted but the contribution decays with the length of time interval, so the probability to visit the forum i is

$$\Pi_i = \frac{\sum_{j=1}^{f_i} \left(\Delta n_j^{(i)}\right)^{-\xi}}{\sum_{h \in \Gamma_u} \sum_{j=1}^{f_h} \left(\Delta n_j^{(i)}\right)^{-\xi}}, \quad (3)$$

where Γ_u is the set of forums being visited by the target user u, and the decaying factor $\xi \approx 1.89$ is estimated from Fig. 6(b). The validation of the individual dynamic model is demonstrated in Fig. 7(b) and 7(c), from which one can see that the numerical results consist well with the empirical scaling behaviors for both expansion and revisit.

IV. CONCLUSION

In this paper, we comprehensively investigate the dynamic patterns of researchers' academic forum activities in Sciencenet. We show a power-law scaling between the frequency of visits to an academic forums and the number of corresponding visitors, which is similar to the allometric scaling law found in biology systems [39, 40]. Meanwhile, at the individual level, the number of distinct visited forums $S(n)$ increases with the number of visits n in a power-law behavior, obeying the well-known Heaps' Law [41, 42], which is also similar to the previous studies on portal browsing activity and human mobility [35, 36]. The memory effect in academic forum activity is unveiled

by the power-law distributions of the real time interval τ and the number of visits Δn taken to revisit the same academic forum.

Inspired by these empirical results and a previous theoretical model [35], we propose a dynamic model, incorporating with the exploration of new academic forums and the preferential return with memory effect. Through extensively experimental testing, the numerical results from this dynamic model agree well with the empirical observations, suggesting its validation. We have also checked that the lack of each of the three ingredients will lead to huge deviation from the real statistics.

[1] A.-L. Barabási, Nature **435**, 207 (2005).
[2] A. Vázquez, Phys. Rev. Lett. **95**, 248701 (2005).
[3] A. Vázquez, J. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Phys. Rev. E **73**, 036127 (2006).
[4] C. Anteneodo, Phys. Rev. E **80**, 041131 (2009).
[5] X.-P. Han, T. Zhou, and B.-H. Wang, New J. Phys. **10**, 073010 (2008).
[6] M.-S. Shang, G.-X. Chen, S.-X. Dai, B.-H. Wang, and T. Zhou, Chin. Phys. Lett. **27**, 048701 (2010).
[7] Z.-D. Zhao, Z. Yang, Z.-K. Zhang, T. Zhou, Z.-G. Huang, and Y.-C. Lai, Sci. Rep. **3**, 3472 (2013).
[8] M. Szell, R. Sinatra, G. Petri, S. Thurner, and V. Latora, Sci. Rep. **2**, 457 (2012).
[9] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, Proc. Natl. Acad. Sci. U. S. A. **102**, 9424 (2005).
[10] A. Vazquez, Physica A **373**, 747 (2007).
[11] K. Goh and A.-L. Barabási, Europhys. Lett. **81**, 48002 (2008).
[12] S.-M. Cai, Z.-Q. Fu, T. Zhou, J. Gu, and P.-L. Zhou, Europhys. Lett. **87**, 68001 (2009).
[13] Z.-D. Zhao, S.-M. Cai, J. Huang, Y. Fu, and T. Zhou, Europhys. Lett. **100**, 48004 (2012).
[14] V. Alfi, G. Parisi, and L. Pietronero, Nat. Phys. **3**, 746 (2007).
[15] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. Schellnhuber, Proc. Natl. Acad. Sci. U. S. A. **107**, 18803 (2010).
[16] J. Stehlé, A. Barrat, and G. Bianconi, Phys. Rev. E **81**, 035101 (2010).
[17] K. Zhao, J. Stehlé, G. Bianconi, and A. Barrat, Phys. Rev. E **83**, 056109 (2011).
[18] K. Zhao, M. Karsai, and G. Bianconi, PLoS ONE **6**, e28116 (2011).
[19] J. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Proc. Natl. Acad. Sci. U. S. A. **104**, 7332 (2007).
[20] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).
[21] L. Lü, D.-B. Chen, and T. Zhou, New J. Phys. **13**, 123005 (2011).
[22] S. Meloni, N. Perra, A. Arenas, S. Gómez, Y. Moreno, and A. Vespignani, Sci. Rep. **1**, 62 (2011).
[23] M. Salganik, P. Dodds, and D. Watts, Science **311**, 854 (2006).
[24] R. Dukas, Brain Behav. Evolut. **63**, 197 (2004).
[25] R. Reis, J. Monet. Econom. **53**, 1761 (2006).
[26] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, Phys. Rep. **519**, 1 (2012).
[27] T. Zhou, R.-Q. Su, R.-R. Liu, L.-L. Jiang, B.-H. Wang, and Y.-C. Zhang, New J. Phys. **11**, 123008 (2009).
[28] R. Guimerà, A. Llorente, E. Moro, and M. Sales-Pardo, PLoS ONE **7**, e44620 (2012).
[29] T. Zhou, H. A. T. Kiet, B. J. Kim, B.-H. Wang, and P. Holme, Europhys. Lett. **82**, 28002 (2008).
[30] R. D. Malmgren, D. B. Stouffer, A. S. L. O. Campanharo, and L. A. N. Amaral, Science **325**, 1696 (2009).
[31] F. Radicchi, Phys. Rev. E **80**, 026118 (2009).
[32] Z.-D. Zhao and T. Zhou, Physica A **391**, 3308 (2012).
[33] H. S. Heaps, *Information retrieval: Computational and theoretical aspects* (Academic Press, Inc., 1978).
[34] Y.-L. Zha, T. Zhou, and C. Zhou, arXiv: 1507.05248 (2015).
[35] C. Song, T. Koren, P. Wang, and A.-L. Barabási, Nat. Phys. **6**, 818 (2010).
[36] A. Chmiel, K. Kowalska, and J. Hołyst, Phys. Rev. E **80**, 066122 (2009).
[37] F. Guo, Z. Yang, Z.-D. Zhao, and T. Zhou, arXiv: 1506.09096 (2015).
[38] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Rev. **51**, 661 (2009).
[39] M. Kleiber, Hilgardia **6**, 315 (1932).
[40] G. B. West, J. H. Brown, and B. J. Enquist, Science **276**, 122 (1997).
[41] L. Lü, Z.-K. Zhang, and T. Zhou, PloS ONE **5**, e14139 (2010).
[42] L. Lü, Z.-K. Zhang, and T. Zhou, Sci. Rep. **3**, 1082 (2013).